INTERPOLATION AND ANALYSIS OF SURFACES

1. Introduction

Spatial Interpolation plays a significant role in obtaining a continuous estimate of variables at unobserved locations based on measured values at known locations. In this project, we focused on the interpolation of ocean bathymetry using point data in the Gulf of Mexico.

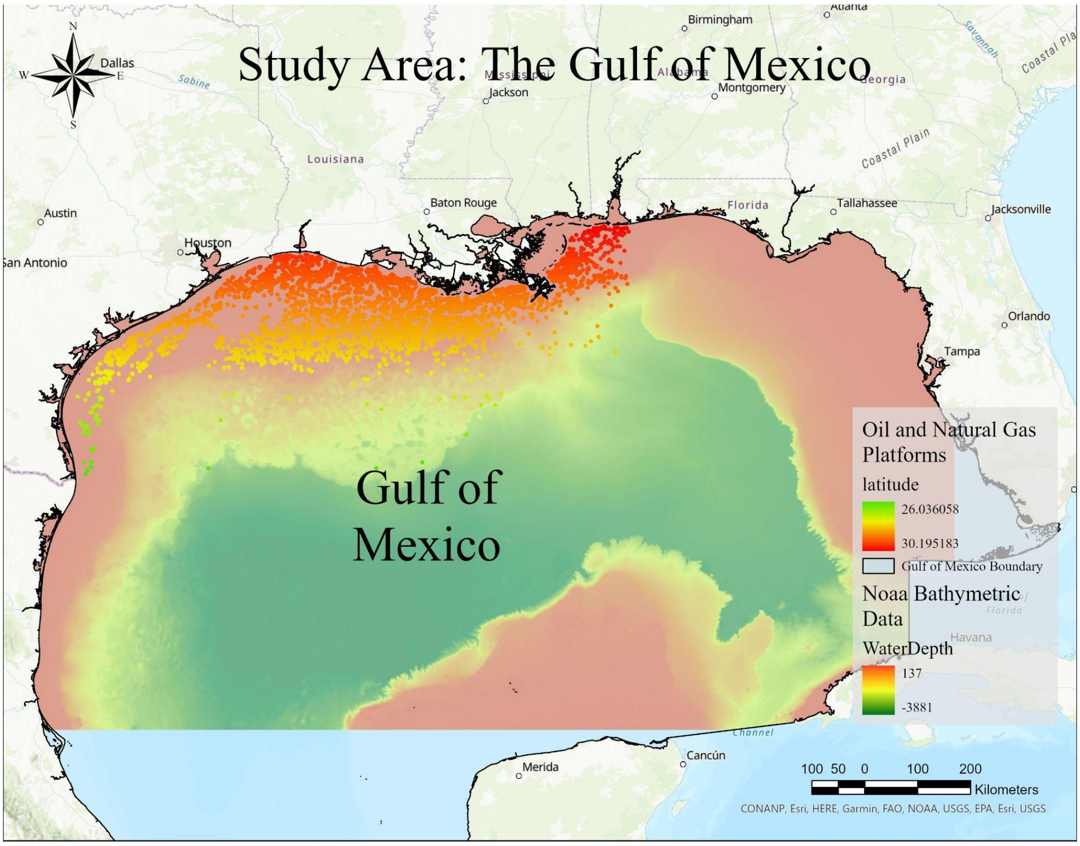


Fig.1 Study Area [1]

As shown in Fig.1, the study area of this project is the Gulf of Mexico, a basin surrounded by the United States, Mexico, and Cuba. It is an essential region for commercial and recreational activities. The bathymetry of the Gulf is complex, with features such as ridges, canyons, and abyssal plains.

In this project, we explored two interpolation techniques, including Inverse Distance Weighting(IDW) and Kriging, and assess their accuracy when compared to continuous bathymetric data from the satellite images. IDW is a deterministic interpolation method that assumes the values at unsampled locations are influenced more by nearby sample points than distant ones. On the other hand, Kriging is a geostatistical interpolation method that uses a spatial correlation model to estimate the values at unsampled locations.

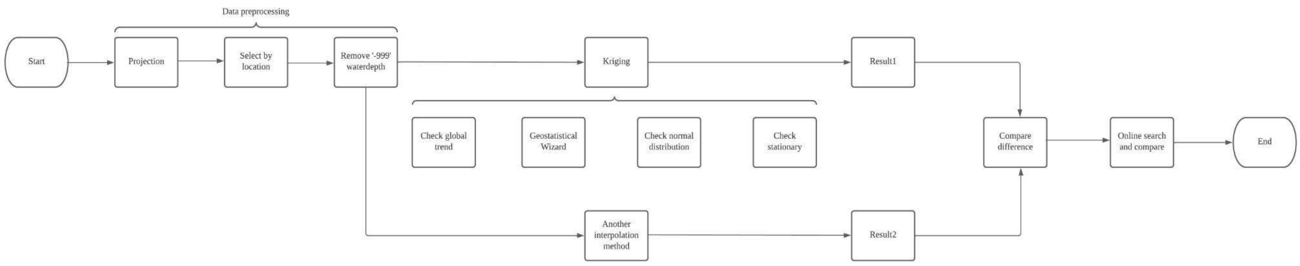


Fig.2 Workflow

Fig.2 is the workflow for this project, involving several steps to interpolate bathymetric data using Kriging and IDW interpolation methods. We started by collecting point data for the Gulf of Mexico bathymetry measurements from the oil and natural dataset. The data preprocessing included projecting the data onto a suitable coordinate system for the study area, cutting the point layer within the study area boundary, and removing missing values. We then performed Kriging and IDW separately. For the Kriging part, we took care to check for global trends and ensure that the interpolated values followed a normal distribution and were stationary. We compared the results of the two methods with the deepwater Gulf of Mexico bathymetry grid.

2. Data and data pre-processing

The data used in this project are listed in Table 1. The oil and natural gas platform dataset is obtained from the Homeland Infrastructure Foundation-Level Data (HIFLD) and contains point data for oil and natural gas drilling platforms located off the coast of the United States. The dataset includes the latitude and longitude coordinates of each platform and provides information on the water depth. The Gulf of Mexico Boundary dataset consists of a line boundary which is derived from the National Oceanic and Atmospheric Administration (NOAA). The Deepwater Gulf of Mexico Bathymetry Grid dataset obtained from the Bureau of Ocean Energy Management is used as the comparison of the results of the interpolation results.

Table.1 Dataset Table

Datasets	**Source**	**Description**
Oil and Natural Gas Platforms	HIFLE [2]	Point data of oil and natural gas drilling platforms on the coast of the US
Gulf of Mexico Boundary	NOAA [3]	Line Boundary of the Gulf of Mexico
Deepwater Gulf of Mexico Bathymetry Grid	Deepwater Gulf of Mexico Bathymetry Grid [4]	Deepwater Gulf of Mexico Bathymetry Grid

The data pre-processing part includes projection, clip, and removing missing values. The data used in this project was projected into the North America Lambert Conformal Conic projection because it has a large range on the map in North America. Then the points were clipped using the Gulf of Mexico boundary. Finally, points with a water depth of "-999" were removed because these points are meaningless for our interpolation analysis.

3. Methods

3.1. kriging

Kriging is a method used to estimate values at unsampled locations based on the spatial correlation of the data. It assumes that the variable being interpolated is a realization of a stochastic process with a known mean and covariance function, and it calculates the weights for the nearby sampled values based on their spatial correlation with the unknown location. These weights are then used to estimate the unknown value at the location of interest. Kriging provides a measure of the uncertainty in the estimated value and can be used to make spatial predictions or create contour maps of the variable being interpolated.

In this project, the goal is to analyze the spatial interpolation in water depth of bathymetric data in the Gulf of Mexico, so a suitable kriging method should be selected. There are several kriging methods in ArcGIS Pro, and the following shows the characteristics of different ways.

Ordinary Kriging is a widely used method when the data are normally distributed and the mean and variance are constant throughout the study area. Simple Kriging is useful when there is a known mean value and the data are normally distributed. Universal Kriging is used when the data have a trend, and the trend is known or can be estimated from other variables. Indicator Kriging is used when the data are categorical and the goal is to estimate the probability of a certain class or category at unsampled locations. Probability Kriging is used when the goal is to estimate the probability distribution of the variable being interpolated. Hence, some pre-inspection should be done like checking data distribution and stationarity to choose the best kriging method.

Firstly, to check the global trend in the dataset, we used the "trend" geoprocessing tool in ArcGIS Pro to get the raster surface figure from point water depth using a trending technique, which shows the water depths do have the obvious global trend from coast to the deep sea as is shown in Fig. 3, the polynomial order is 1 this time. Also, the number of polynomial orders could be increased to more like 5.

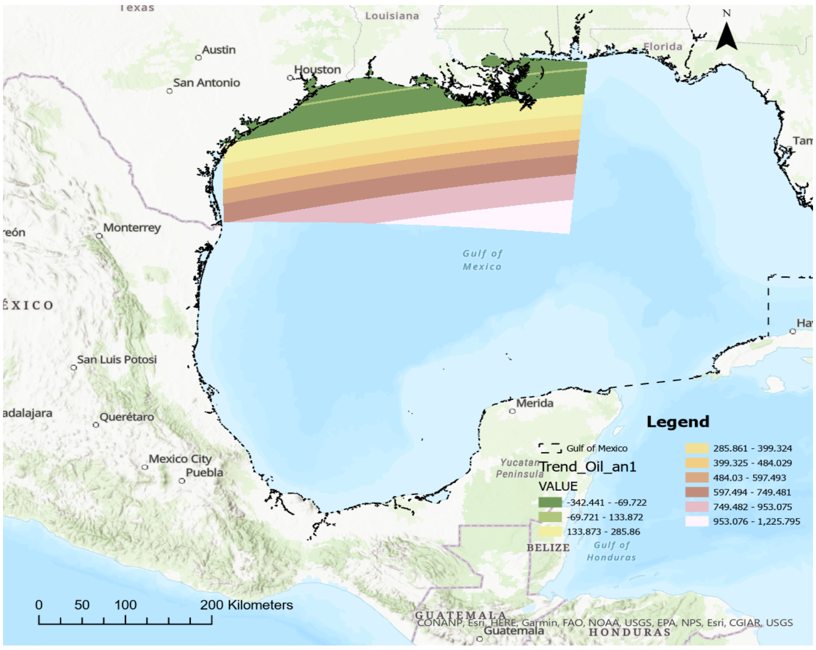


Fig.3 Trend of Water Depth

And, Geostatistical Wizard shown in Figure could help find whether there is a trend or not, when I selected the “Global Polynomial Interpolation” and polynomial order as 5 because it would have the lowest Root-Mean-Square Error as is shown in Fig. 4, which indicates the trend is very obvious.

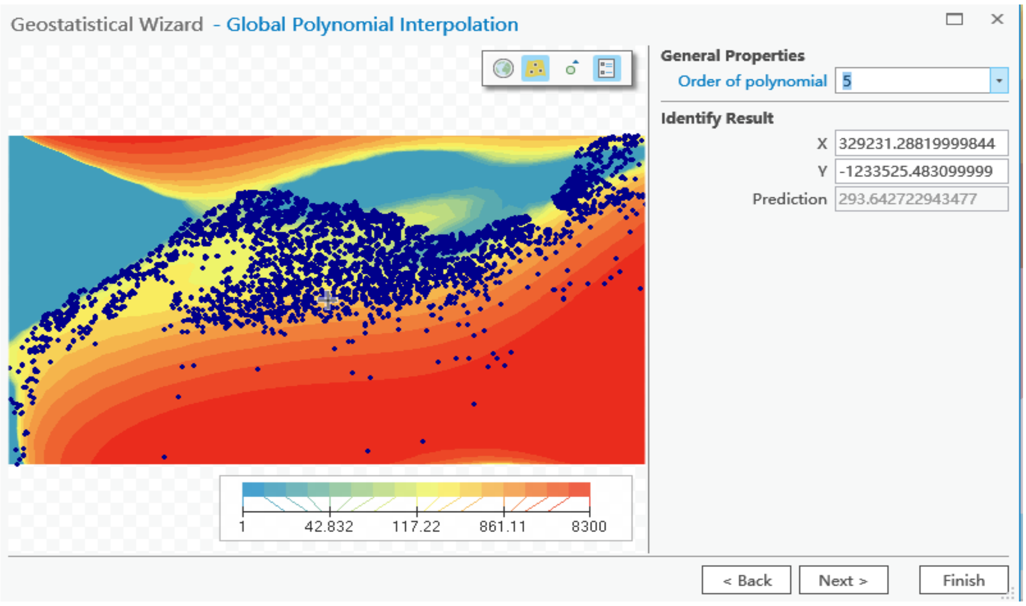


Fig.4 Global Polynomial Interpolation

Besides that, the scatter plot shown in Fig. 5 could show the relationship between water depth and latitude, the figure shows that the water depth would be deeper when latitude decreases, which means the dataset has a significant global trend.

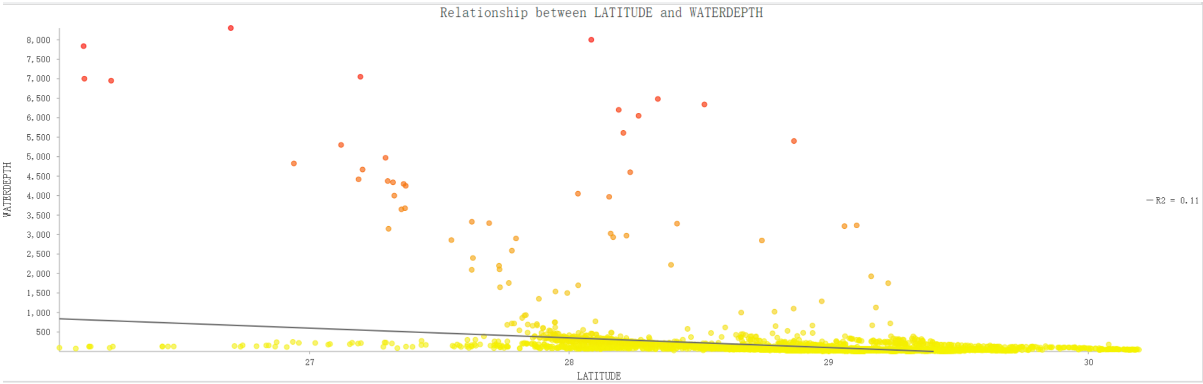


Fig.5 Scatter Plot Between Latitude and Water Depth

Therefore, there is an obvious correlation between water depth and latitude, global trend must be considered in the Kriging method.

Next, it is necessary to check whether the data is normally distributed. The histogram in Fig. 6 shows the distribution of water depth in the ocean, and the data have been transformed logarithmically to find out if it indicates the normal distribution. The QQ plot in Fig. 7 also shows the relationship between water depth and normal distributions, they have good fitting mostly.

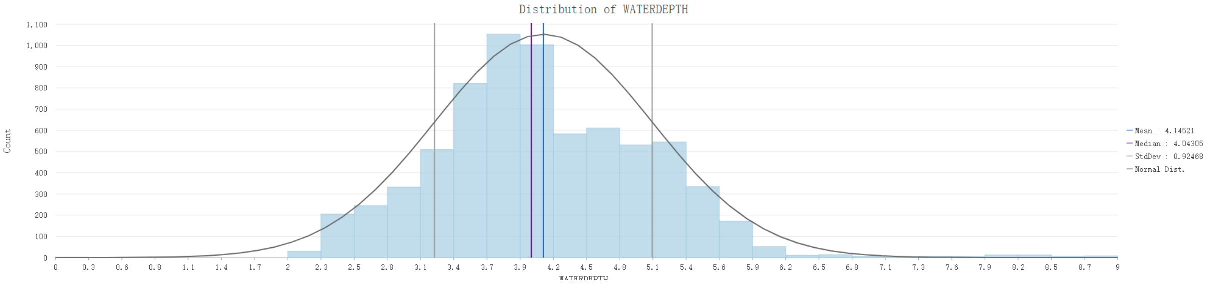


Fig.6 Water Depth Distribution

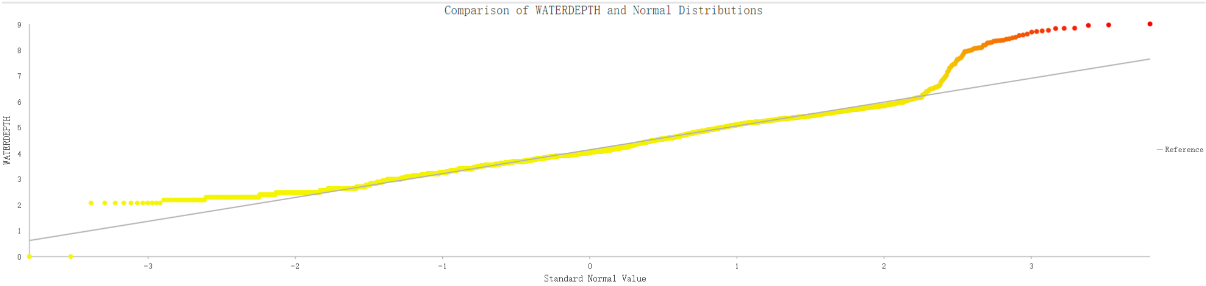


Fig.7 Water Depth Scatter Plot

After that, the “subset features” geoprocessing tool helps select 30% of data from the original dataset randomly three times to get the mean and standard deviation of sub-regions of the data, which is found the two subset distributions are similar as shown in Fig. 8, Fig. 9 and Fig. 10, so data stationery is not a problem.

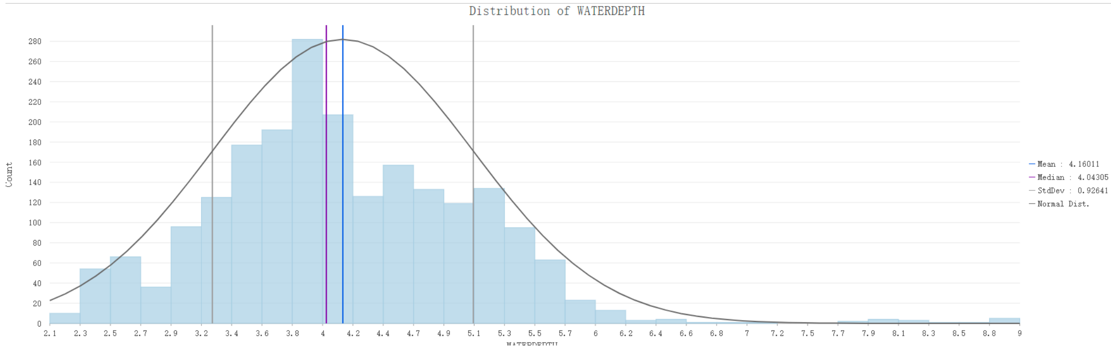


Fig.8 Sub-region 1

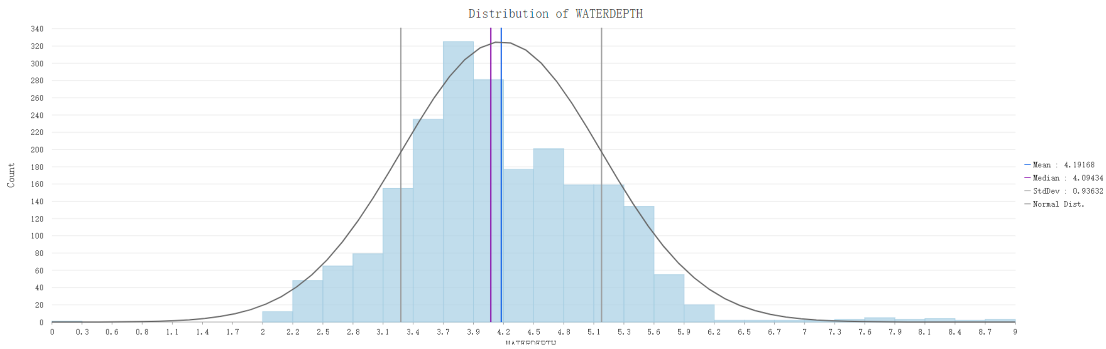


Fig.9 Sub-region 2

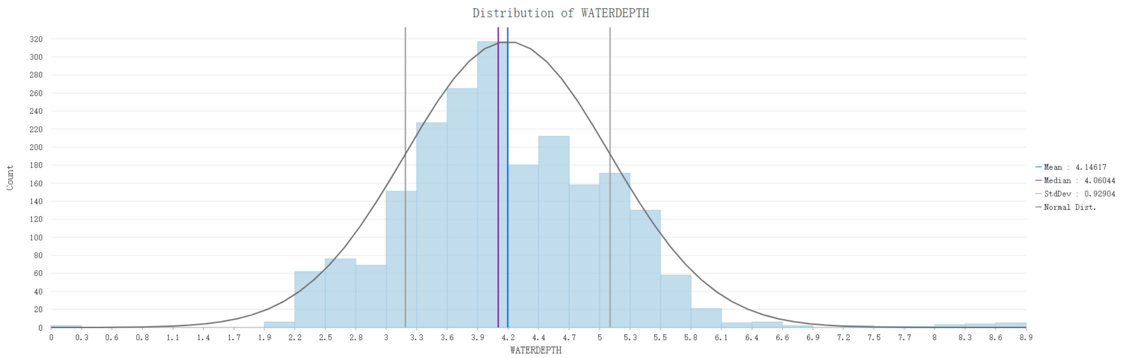


Fig.10 Sub-region 3

Hence, the water depth data are normally distributed, then Ordinary Kriging or Simple Kriging may be suitable. And the dataset does not have other spatially correlated variables, such as temperature or salinity. And the project does not focus on estimating the probability of exceeding a certain water depth threshold or identifying areas with a high probability of extreme water depths.

Moreover, the stationary of different regions are similar, and Simple Kriging assumes that the mean and variance of the variable being interpolated are constant throughout the study area, whereas Ordinary Kriging assumes that the mean is constant but the variance may vary spatially. As the variance in sub-regions is so similar, the ordinary kriging method is chosen in the end.

As shown in Fig. 11, with some pre-processing for the dataset, like in the first step, it is found that the dataset could be likely normally distributed if the values are transformed logarithmically, so the transformation type is “log”. To remove the global trend in the dataset and keep the model simpler and not overfitted, the order of trend removal is “first”. And the goal is to estimate the prediction of water depth in the Gulf of Mexico, so “Prediction” was noted.

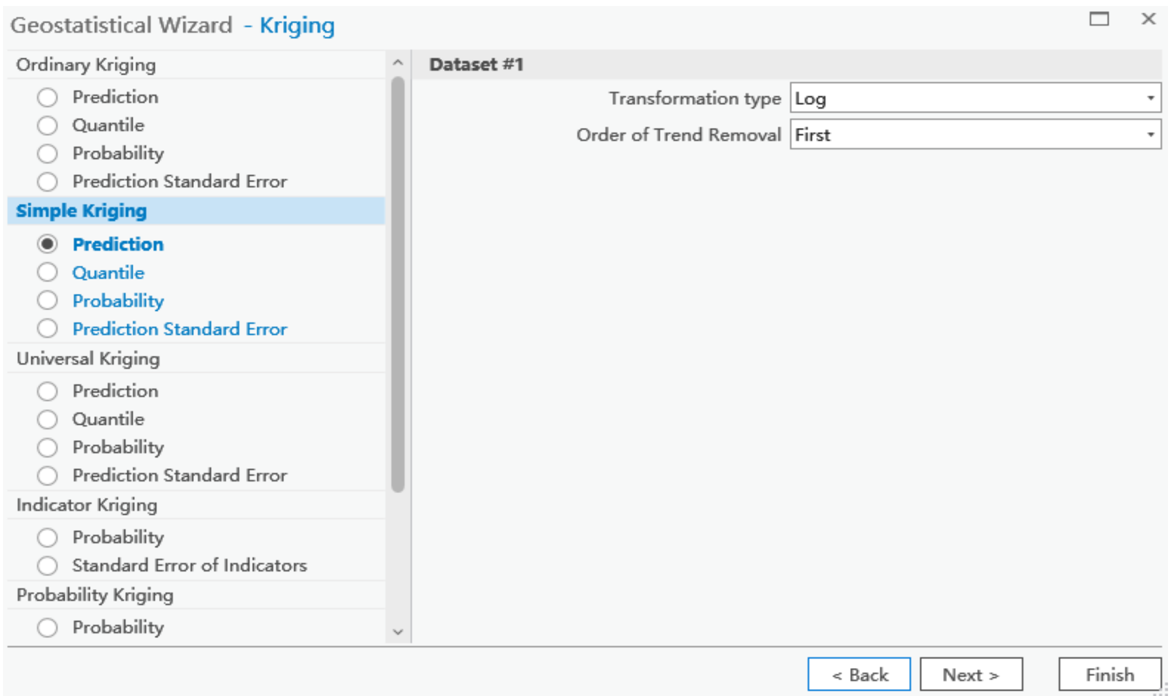


Fig.11 Simple Kriging Dataset

Next, the geostatistical wizard generates a semivariogram with blue crosses showing the average variation for each pair of points.

It is seen in the semivariogram plot on the left by selecting the Function Type in Fig. 12. A goal with kriging is to choose a model that "fits" the scatter of points. There are several models from which to choose. These models are found under the pull-down menu located to the right of the Model #1 option. Choosing different models automatically updates the "fit" (or blue line) in the semivariogram.

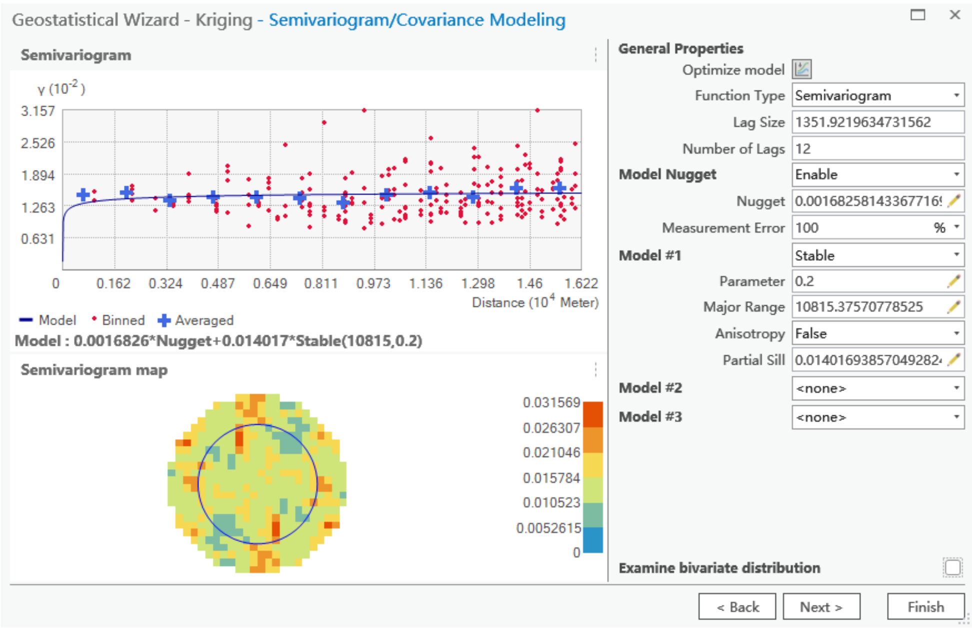


Fig.12 Semivariogram Diagram

The lag size is the size of a distance class into which pairs of locations are grouped. As a rule of thumb, people can multiply the lag size by the number of lags for it to equal half of the largest distance among all points. If points aren't clustered, people can run the 'Average Nearest Neighbor' tool which tells the average distance between points.

Generally, ArcGIS Pro has added the functionality to optimize all these parameters for people. While clicking the optimize button, it will find the value for each parameter that results in the smallest root mean square error. That would be a lot of trial-and-error for the user to test each scenario. Ultimately, it's usually best to go with the semivariogram model that the software considers to be the best.

After getting the optimized model, it can get previews of what the interpolated surface will look like given the currently selected parameters. Also, specify how many neighbors to include using the Neighborhood Type options. People can specify how many Neighbors to Include in the local estimates and also how they are distributed around the location to be estimated using the Sector Type options. The various 'pie slice' or sector options define several regions around the estimation location. These sectors are each required to contain the defined number of neighboring control points, as specified by the minimum and maximum neighbors options. In this project, each parameter is the default. The parameter selections are shown in Fig. 13.

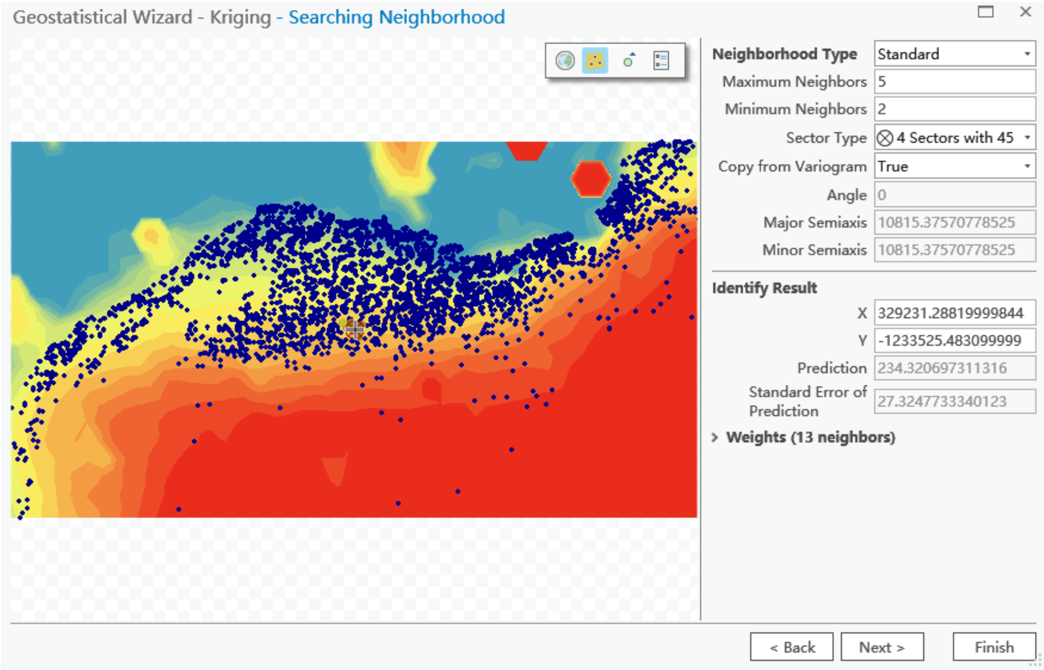


Fig.13 Searching Neighborhood

Finally, The cross-validation step for kriging takes one of the input data points and throws it out of the data set. It runs the prediction back to that location using all the remaining points. Again, people know what the true value is, this process uses all remaining to predict that value. For cross-validation, it iterates through all of the input points until it's complete. Then, it creates this summary table of residuals comparing actual versus the model's predicted values. What this table shows is how robust your model is.

To know the performance of the model and how close the estimated data is to the real one as shown in Fig. 14. To put these all in perspective, root-mean-square standardized, as it should be close to 1. In addition, the root mean square error should be as small as possible.

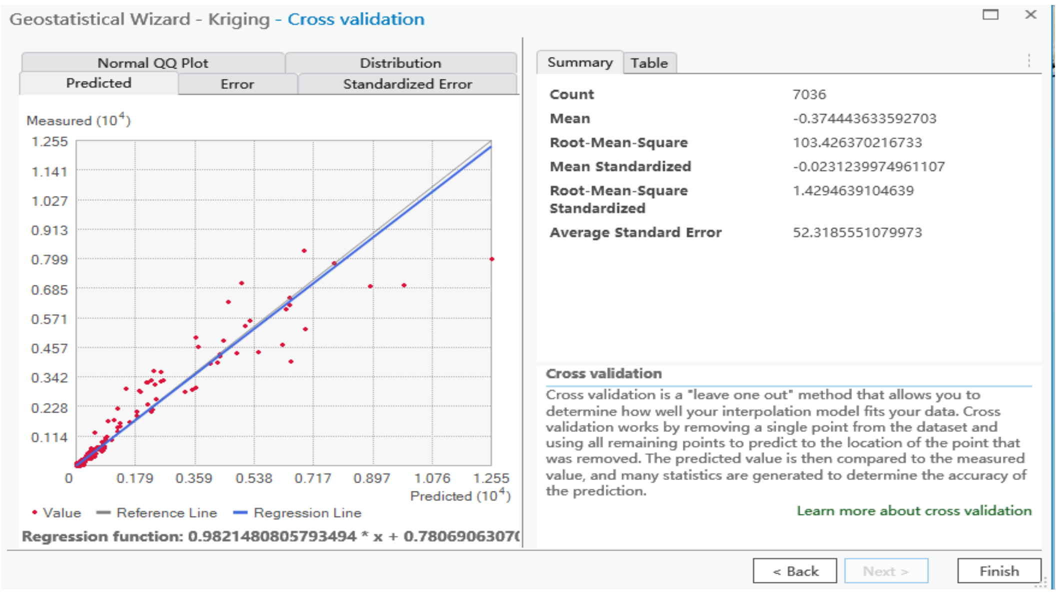


Fig.14 Cross Validation and Model Summary

In the end, the predicted interpolation dataset could be exported and transferred into raster data.

3.2. IDW or kernel interpolation

Inverse Distance Weighted (IDW) interpolation is a deterministic spatial interpolation approach, which is introduced to estimate an unknown value at a location using some known values with corresponding weighted values. Spatial autocorrelation is its underlying assumption, which is a phenomenon where the values of a variable at one location are correlated with the values of the same variable at nearby locations.

To realize this method for estimating the water depth overall study areas, this project uses IDW in ArcGIS Pro. It determines cell values using a linearly weighted combination of a set of sample points. The weight is a function of inverse distance. The parameters we set for this geoprocessing tool are shown in Table 2. The depth_contours_500m_intervals is the multi-polygon feature showing the water depth contour at 500, 1000, 1500, and 2000 meters. (Fig15)

Table.2 Parameters Decision for IDW

Parameter	**Value**	**Description**
Z value field	WATER DEPTH	The surface being interpolated
output cell size	default: 1903	The cell size of the output raster that will be created.
Power	default: 2	The exponent of distance
Number of Points	default: 12	The number of nearest input sample points to be used to perform the interpolation.
Input barrier polyline features	depth_contours_500m_intervals [5]	Polyline features to be used as a break or limit

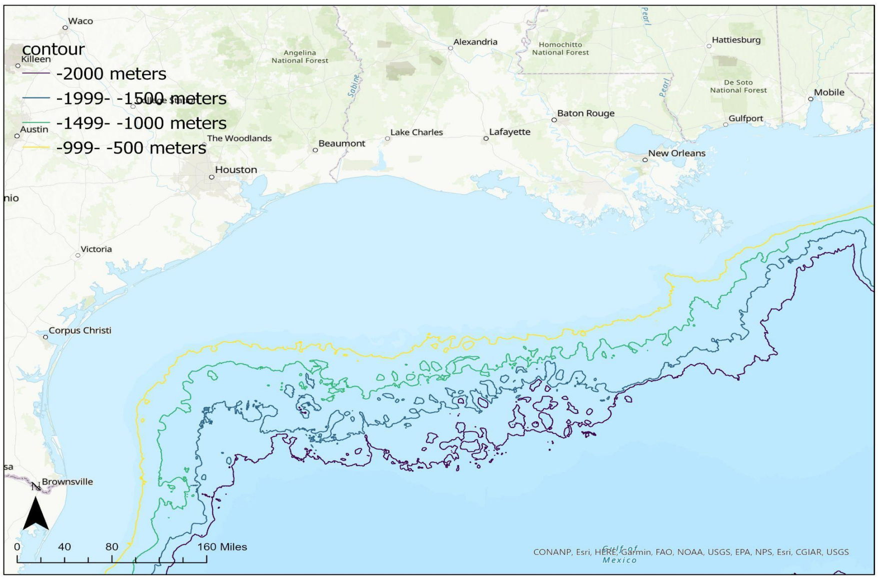


Fig.15 Depth Contours by 500m Intervals

4. Results

4-1 Result of Kriging

The figure shows the output of the simple kriging method interpolation results for water depth in the Gulf of Mexico, the lighter colors like yellow and orange indicate the smaller depth and the deeper red means the larger water depth. We can observe from this map that the seawater depth in the Gulf of Mexico becomes shallower as it approaches the northern inland area, while it becomes deeper towards the southeast with clear stratification. Overall, the water level varies significantly across the map. The depth variation is more concentrated towards the northern coast, but there is still significant depth variation in other areas. The resulting map shows a large red area in the southeast, indicating that the Gulf of Mexico is deeper and the seabed is flatter in that area. This result agrees with the actual ocean structure, where continental shelves and slopes are steeply sloped, while the seafloor is relatively flat.

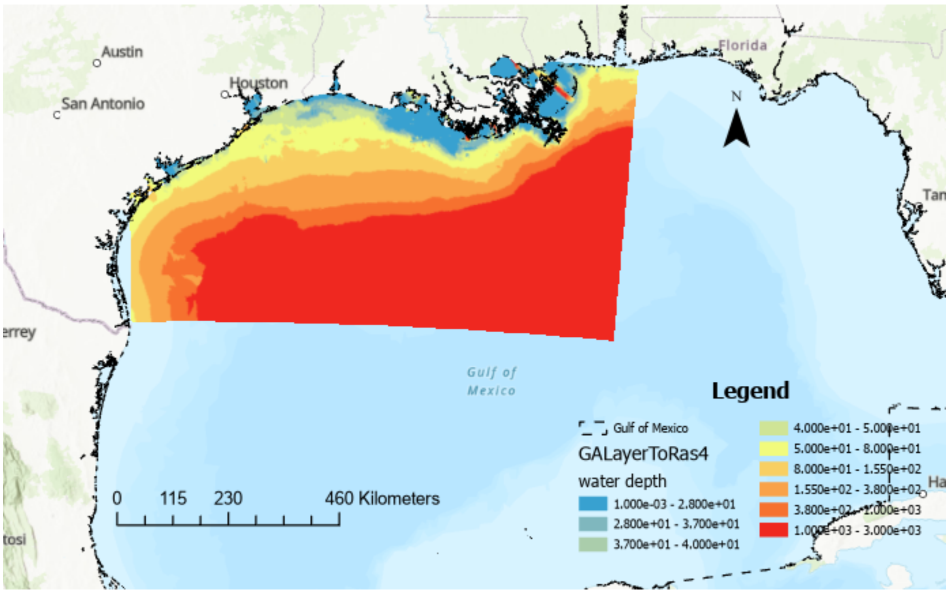


Fig.16 Simple Kriging Interpolation Result

4-2 Result of IDW

As the result of IDW interpolation, we generated a raster layer called Idw with colors representing the value of predicted water depth. In this map, the lighter colors like green and yellow mean the smaller depth of the water while the deeper colors like red indicate a greater depth. The information presented in the IDW result map is largely consistent with the Kriging method. Both methods show that water depth decreases from the northwest to the southeast and exhibits clear stratification. However, the IDW result covers a larger area of the Gulf of Mexico than the Kriging method, likely due to the different interpolation parameters used during the analysis.

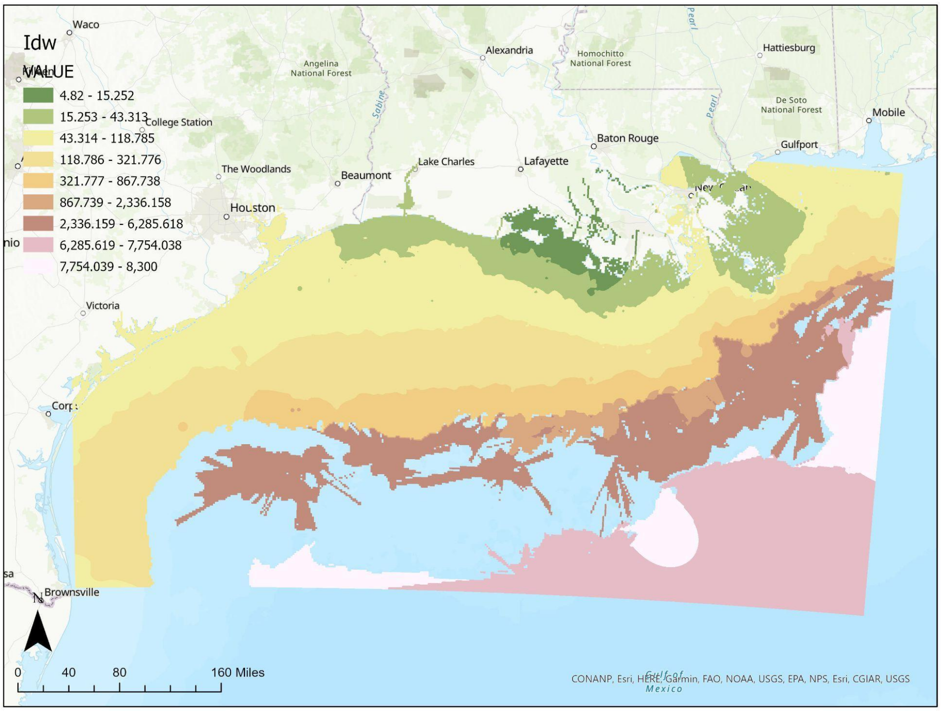


Fig.17 IDW Interpolation Result

4-3 Comparison between two methods and bathymetry data

To strengthen the credibility of our findings, we compared the visual output of the Kriging and IDW interpolation methods with the actual data of the Northern Gulf of Mexico Deepwater Bathymetry Grid obtained from 3D Seismic. We performed this visual comparison in ArcGIS Pro to analyze our results.

We have opted to utilize BOEM's deepwater Gulf of Mexico bathymetry grid as the existing bathymetry dataset [3]. This comprehensive dataset is composed of more than 100 3D seismic surveys that have been merged. The XY pixel size is 40 feet, and it provides water depth measurements in both feet and meters. The average depth error is just 1.3%, with the maximum error occurring in shallow water depths (less than 200 feet) of the outer continental shelf. The dataset is already established with WGS 1984 Web Mercator (auxiliary sphere) projected coordinate system and WGS 1984 geographic coordinate system, and thus requires no preprocessing for our purposes. The only thing we do is reclassify this data into 9 classes(Fig.18). As we can see clearly in this histogram, the water depth range in real data is from -130 feet to -11087 feet. Besides, the distribution of seawater depth is non-uniform, and there is a large range of seawater in the middle depth area. This area’s seawater depth varies greatly, indicating a steep seabed slope.

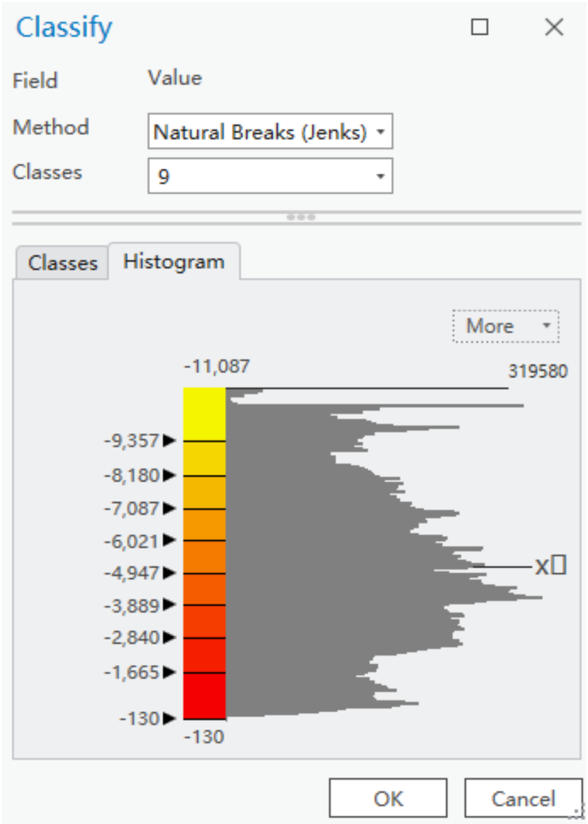


Fig.18 Reclassification of real data

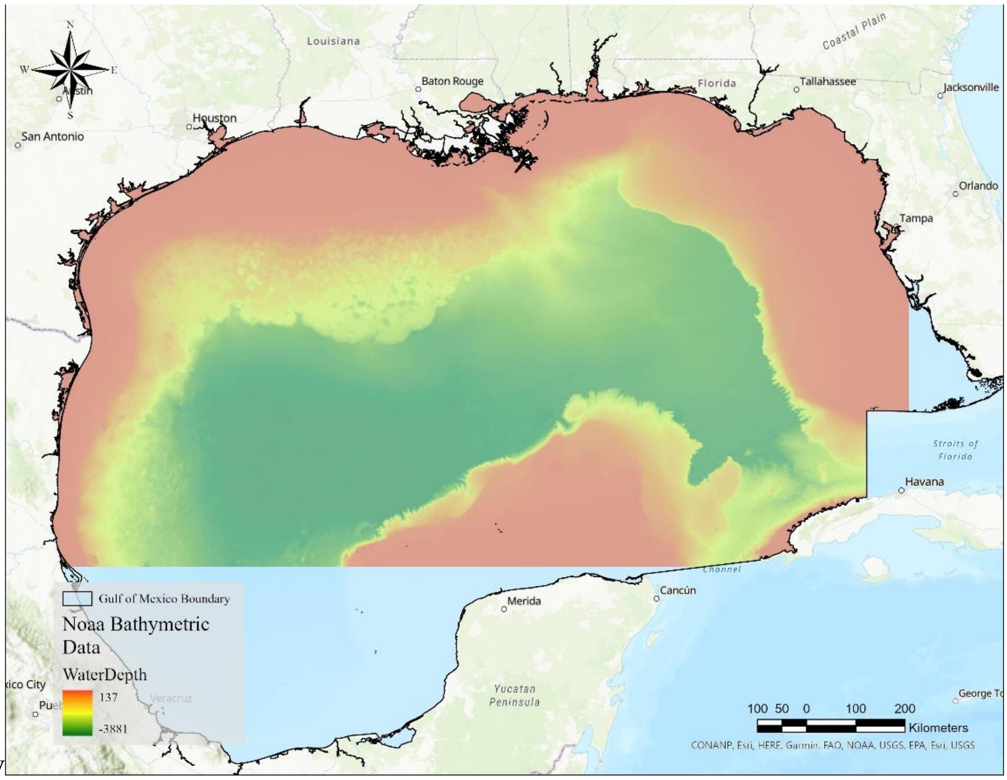
eweeeeewew 

Fig.19 Real Bathymetry

The results from the Kriging method show that the depth of the sea in the southern area is deeper, which is consistent with the actual data. Additionally, the depth levels depicted in both the Kriging result map and the real data map are similar.

Visual Comparison with IDW result:

Upon comparing the result of IDW with the actual data, we observed that the difference in water depth levels between the two maps is consistent. Both maps exhibit the same trend where the sea depth gradually decreases as we move towards the northern part of the US land and deepens towards the southeast. Furthermore, we observed that both maps present a similar range of water depths. The actual data displays a range of water levels between -130 feet and -11087 feet. In the IDW-generated maps, the water depths range from -482 feet to -8803 feet.

Overall, despite some differences in the extent of the covered seawater and specific water depths, the results obtained by the Kriging and IDW methods generally reflect the characteristics and real data of the seawater depth in the Gulf of Mexico. Both methods show the main features of the region with a relatively small water depth in the northwest and a relatively large water depth in the southeast. Additionally, they both indicate that the seawater depth in the bay exhibits obvious gradient changes, with a relatively steep slope, consistent with the real data.

5. Discussion

5.1. limitation in kriging interpolation

There also are some limitations in this project, for example the data stationary check process is not standard, select the data points randomly may be likely to get a similar result from the original dataset due to the uncertainty like percentage and the dataset size, another reasonable method is to check your data's stationarity with a Voronoi map symbolizing by entropy (variation between neighbors) or standard deviation and look for randomness. As shown in the Figure, there are three clustering in the map with different mean values of water depth, so it is better to run kriging on sub-regions of the data, and later combine the results into one surface, which may decrease the root mean square value in our model.

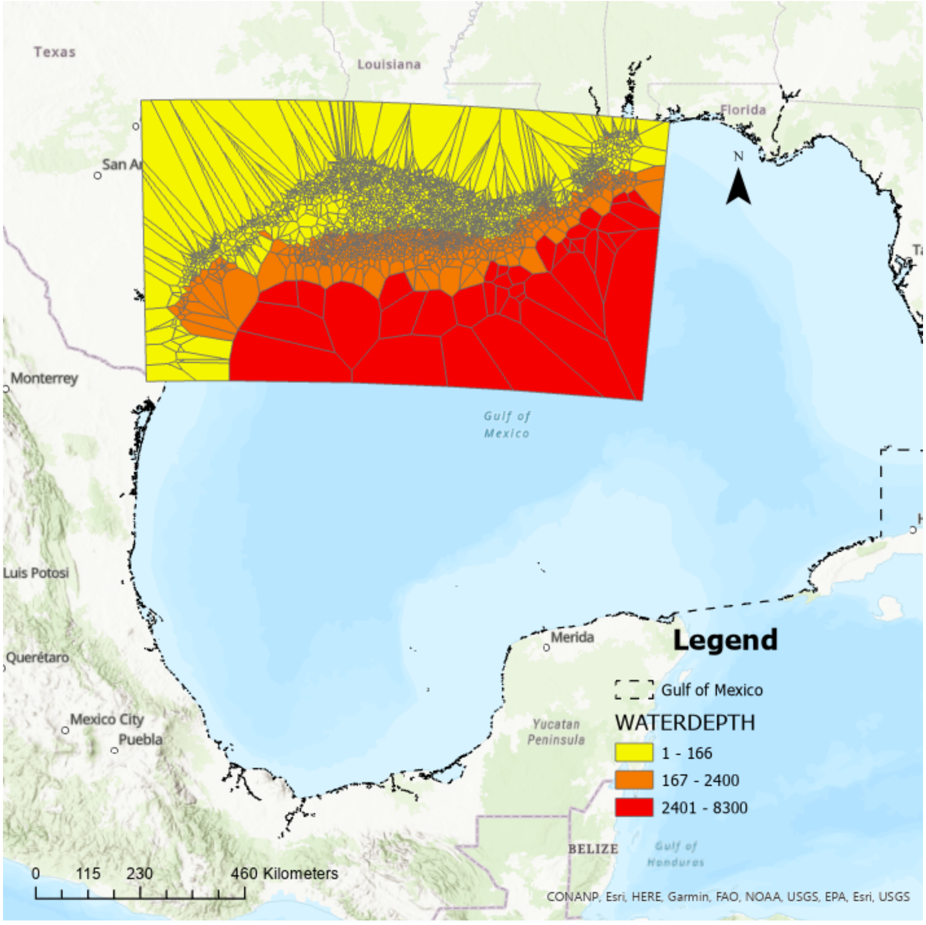


Fig.20 Voronoi Map of Water Depth

A further limitation of our analysis is the relatively small coverage of the real data of the Gulf of Mexico sea depth we utilized. While we were able to visually compare the results of the Kriging and IDW methods in ArcGIS Pro and depict the primary features of the area, a more comprehensive comparison of the final results would be facilitated if we could access depth data with greater coverage of the Gulf of Mexico.

5.2. Why not Kernel

First and foremost, we believe that kernel interpolation is not the best method for this dataset. The biggest issue it faces is the inability to make continuous predictions due to the excessive dispersion of data points.

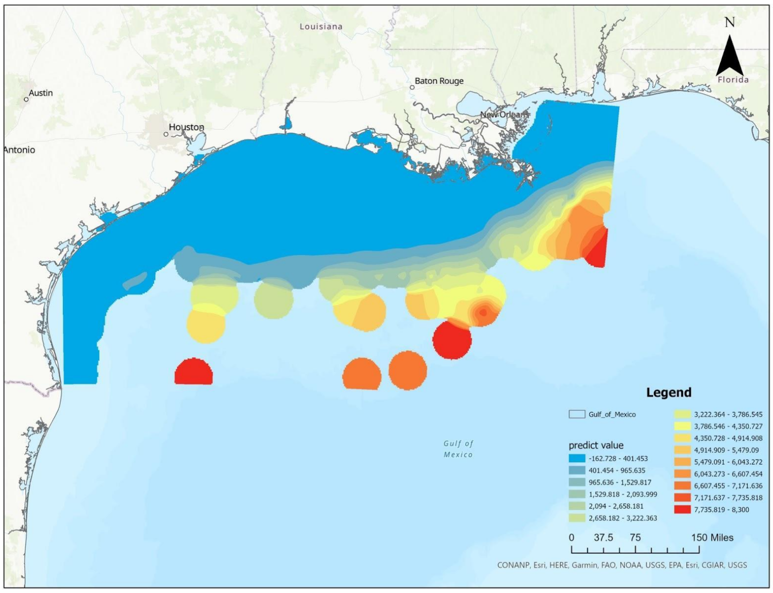


Fig.21 Kernel Interpolation Results

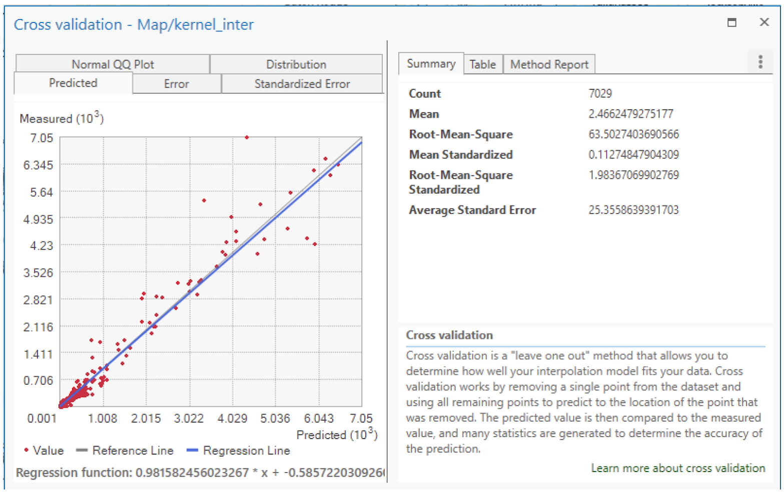


Fig.22 Kernel Interpolation Cross Validation

Kernel interpolation can indeed produce circular patterns around known data points when generating smooth spatial distributions. This phenomenon may be due to the shape of the kernel function and the chosen bandwidth. Kernel interpolation is a local interpolation method, and the interpolation result is influenced by the weight distribution of the kernel function. Kernel function weights are typically highest at the center and decrease with increasing distance. When the bandwidth is set too small, the weight distribution may be overly concentrated, causing the interpolation result to form circular patterns around known data points.

To address this issue, we tried to adjust the parameters of kernel interpolation, including increasing the bandwidth, changing the kernel function, and considering the spatial distribution of data. By increasing the bandwidth, we expand the spatial influence of the kernel function, making the interpolation result smoother. However, too large a bandwidth may cause the interpolation result to be too smooth, ignoring local variations. The choice of kernel function will affect the interpolation result. We tried using different kernel functions (such as Gaussian kernel, Epanechnikov kernel, etc.) to observe their impact on the interpolation result. However, the results changed little. If data points are unevenly distributed in space, it may cause anomalies in the interpolation result. In this case, we can try preprocessing the data, such as using spatial averaging or clustering algorithms to adjust the spatial distribution of the data. This method I believe is the most helpful. However, due to the limit of the results, we are unable to add more data in the southern parts.

6. References

[1] Wikipedia Contributors. "Gulf of Mexico." *Wikipedia*, Wikimedia Foundation, 24 Sept. 2019, en.wikipedia.org/wiki/Gulf_of_Mexico.

[2] "Oil and Natural Gas Platforms." *Hifld-Geoplatform.opendata.arcgis.com*, hifld-geoplatform.opendata.arcgis.com/datasets/oil-and-natural-gas-platforms/explore. Accessed 3 May 2023.

[3] "Marine Regions · Gulf of Mexico (IHO Sea Area)." Www.marineregions.org, www.marineregions.org/gazetteer.php?p=details&id=4288.

[4] "Northern GoM Deepwater Bathymetry Grid from 3D Seismic | Bureau of Ocean Energy Management." *Www.boem.gov*, www.boem.gov/oil-gas-energy/mapping-and-data/map-gallery/northern-gom-deepwater-bathymetry-grid-3d-seismic.

[5] "Depth Contour at Multiple Interval of GEBCO 2020 Bathymetry - Gulf of Mexico (GCOOS)." Gisdata.gcoos.org, gisdata.gcoos.org/datasets/2280ed030cc745e7839cca3faaf98acb/explore?layer=2&location=24.324902%2C-89.130645%2C6.78. Accessed 3 May 2023.